Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

High-throughput Screening for Methionyl-tRNA Synthetases that Enable Residue-specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells

Tae Hyeon Yoo and David A. Tirrell*

Division of Chemistry and Chemical Engineering

Joseph J. Jacobs Institute for Molecular Engineering for Medicine

California Institute of Technology

1200 E. California Boulevard

MC 210-41

Pasadena, CA 91125-4100, USA

*To whom correspondence should be addressed:

David A. Tirrell

Telephone (626) 395-3140

Fax (626) 793-8472

email: tirrell@caltech.edu

Oligonucleotides used in this study

Out_F: 5’-CTTTCGTCTTCACCTCGAG-3’

Out_R: 5’-CTCCATTTTAGCTTCCTTAGCTC-3’

GFR_M78_F: 5’-CTTTGCGCGTTATCCGGATCATnnhAAACGGCATGACTTTTTCAAG-3’

GFP_M78_R: 5’-CTTGAAAAAGTCATGCCGTTTdnnATGATCCGGATAACGCGCAAAG-3’

GFP_M88_F: 5’-GACTTTTTCAAGAGTGCCnnhCCCGAAGGTTATGTACAG-3’

GFP_M88_R: 5’-CTGTACATAACCTTCGGGdnnGGCACTCTTGAAAAAGTC-3’

GFP_M218_F: 5’-CAACGAAATGCGTGACCACnnhGTCCTTCATGAGTTTGTAAC-3’

GFP_M218_R: 5’-GTTACAAACTCATGAAGGACdnnGTGGTCACGCATTTCGTTG-3’

GFP_M233_F: 5’-CTGCTGGGATTACACATGGCnnhGATGAGCTCTACAAATAGAAG-3’

GFP_M233_R: 5’-CTTCTATTTGTAGAGCTCATCdnnGCCATGTGTAATCCCAGCAG-3’

Lib_OutF: 5’-CTTCCTGGCATCTTCCAGGAAATCTC-3’

Lib_R: 5’-AGCGCCTTGAACGGATTCACTTTGTG-3’

MetRS_L13_F: 5’-CTGGTGACGTGCGCAnnkCCGTACGCTAACGGCTC-3’

MetRS_L13_R: 5’-GAGCCGTTAGCGTACGGmnnTGCGCACGTCACCAG-3’

MetRS_Mid_F: 5’-CTACGTCTGGCTGGACnnknnkATTGGCnnkATGGGTTCTTTCAAG-3’

MetRS_Mid_R: 5’-CTTGAAAGAACCCATmnnGCCAATmnnmnnGTCCAGCCAGACGTAG-3’

MetRS_H301_F: 5’-GATATTGTTTACTTCnnkAGCCTGTTCTGGCCTG-3’

MetRS_H301_R: 5’-CAGGCCAGAACAGGCTmnnGAAGTAAACAATATC-3’

MetRS_NheI_F: 5`-TTCCGCgctagcTCTAGAGACGTCCGGCCGGAGCTC-3’

MetRS_NheI_R: 5`-TTTGGGgctagcTCTAGAGACGTCCGGCCGGGTAC-3’

MetRS_SalI_R: 5’- TTTGGGgtcgacTCTAGAGACGTCCGGCCGGGTAC-3’

MetRS_N-terminal_BamHI_F: 5’-TTCCGCggatccATGACTCAAGTCGCGAAGAAAATTC-3’

MetRS_N-terminal_SalI_R: 5'-TTTGGGgtcgacTCATTTAGAGGCTTCCACCAGTG-3'

GFP library construction

Three PCR reactions were performed with PfuUltra™ High-Fidelity DNA Polymerase (Strategene) using pQE-80L/L024_3-3 as a template and the following pairs of primers: Out_F and GFP_M78_R, GFP_M78_F and GFP_M218_R, and GFR_M218_F and Out_R. The DNA fragments obtained from these PCR steps were purified on a 1.5 % agarose gel (QIAquick gel extraction kit, QIAgen). Equimolar quantities of the fragments were mixed and assembled by PCR, and the reaction mixture was purified using Zymo-spin columns (Zymo Research, Orange, CA). The PCR product was amplified using primers, Out_F and Out_R. In order to randomize positions 88 and 233, three PCR reactions were performed using the GFP gene (previously randomized at positions 78 and 218 with NNH codons) as a template and the following pairs of primers: Out_F and GFP_M88_R, GFP_M88_F and GFP_M233_R, and GFP_M233_F and Out_R. The DNA fragments were purified and assembled by PCR. The PCR fragment was digested with *Bam*HI and *Hind*III and ligated into pQE-80L digested with the same enzymes. A 5 μg portion of digested pQE-80L was used for the ligation; the molar ratio of insert to vector was 3. The ligation mixture was purified using Zymo-spin columns and transformed into electrocompetent *E. coli* DH10B cells, yielding more than 10^7 transformants.

MetRS library construction

Four PCR reactions were performed using pMTY11 as template and the following pairs of primers: Out_OutF and MetRS_L13_R, MetRS_L13_F and MetRS_Mid_R, MetRS_Mid_F and MetRS_H301_R, and MetRS_Mid_F and Lib_R. The DNA fragments obtained from these PCR steps were purified on a 1.5 % agarose gel and assembled by PCR. The resulting PCR fragment was digested with *Not*I and *Bsr*GI and ligated into pMTY11 digested with the same enzymes. A 7 μg sample of digested pQE-80L was used for the ligation; the molar ratio of insert to vector was

3. The ligation mixture was purified using Zymo-spin columns and transformed into electrocompetent *E. coli* DH10B cells, yielding more than 5×10^8 transformants. Plasmid DNA library was retransformed into cells of *E. coli* strain DH10B(Met$^-$) bearing the reporter plasmid pQE-80L/GFPrm_AM.

Amino acid activation assays

The expression plasmid for the mutant MetRS (pMTY21) was constructed by ligation between a *Bam*HI/*Sal*I fragment of pQE-80L and the mutant MetRS gene amplified from clone M02c_2-8 using primers MetRS_N-terminal_BamHI_F and MetRS_N-terminal_SalI_R. The form of MetRS truncated at position 548 was used in these experiments. *E. coli* strain DH10B transformed with pMTY21 was grown at 37°C to OD_{600} = 1 in 1 L of 2xYT medium. After induction with 1 mM IPTG, cells were further grown at 25°C overnight. The cells were harvested by centrifugation (6,000 x *g* for 15 min). The mutant MetRS was purified using Ni-NTA chromatography under native conditions according to the manufacturer's instructions. The column eluent was buffer-exchanged with 100 mM Tris buffer (pH 7.5) containing 2 mM DTT by using PD-10 columns (GE Heathcare). The protein solution was mixed with an equal volume of glycerol and frozen at -80°C until needed. The protein concentration was determined by measuring absorbance at 280 nm, assuming an extinction coefficient of the mutant MetRS of 93280 $M^{-1}cm^{-1}$. Activation assays were carried out as described (K. L. Kiick, R. Weberskirch, D. A. Tirrell, *FEBS Lett.* **2001**, *502*, 25-30). The analog concentrations tested ranged from 312.5 μM to 10 mM, and the mutant MetRS was added to reactions at a concentration of 100 nM for Met and 400 nM for Tfn. Data were fit to the Michaelis–Menten model by using Origin software (Origin Lab).

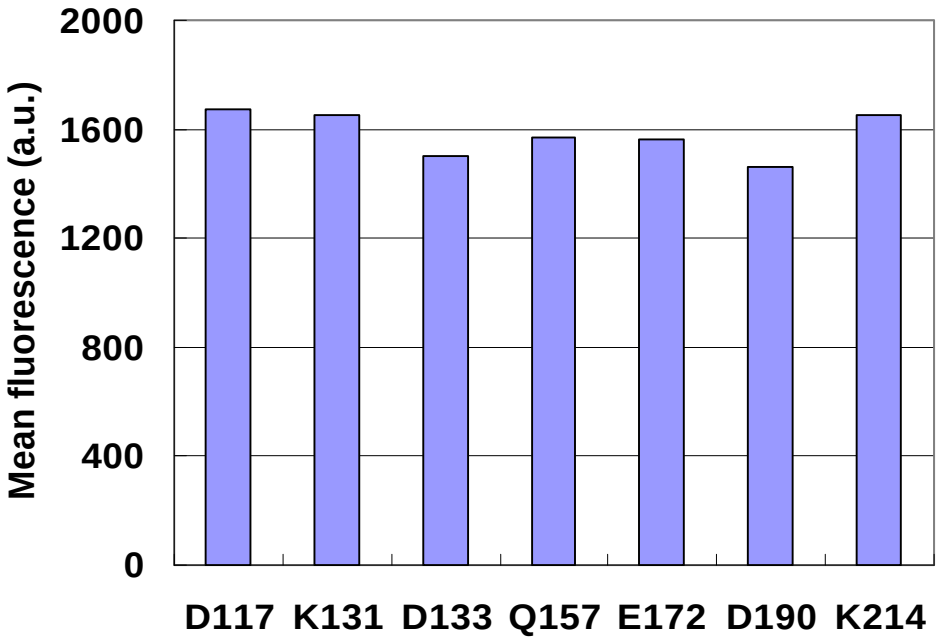

Figure S1. Fluorescence of cells expressing GFPrm libraries in which the indicated position was randomized using an NNK codon. Mean cell fluorescence was measured on a cell sorter equipped with an argon ion laser emitting at 488 nm and a 530/40 bandpass filter.

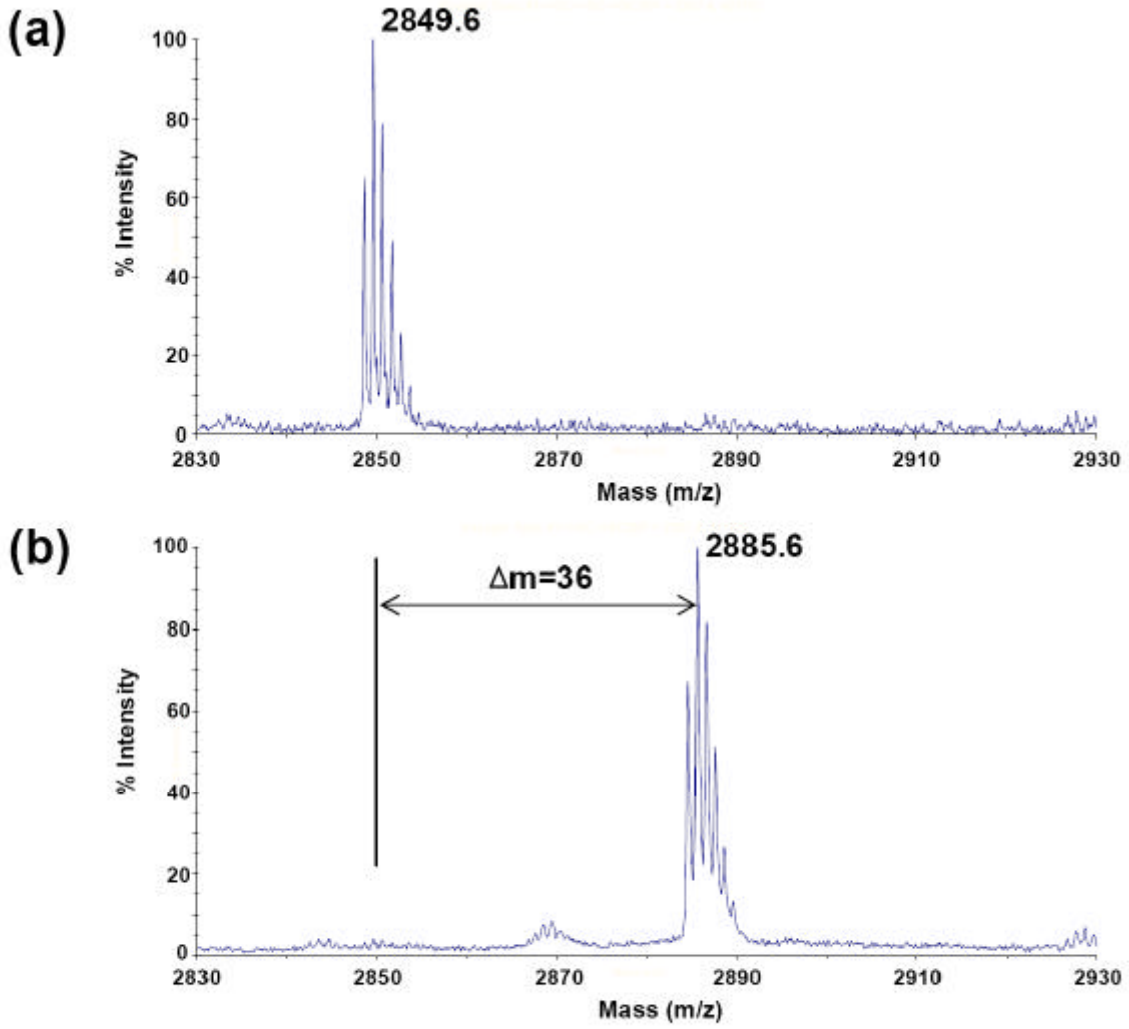

Figure S2. MALDI-MS analysis of GFPrm_AM after trypsin digestion. The protein was expressed with 20 amino acids (a) or 19 amino acids plus 2 mM Tfn (b), and purified under denaturing conditions. A peptide fragment of sequence HNVMDGSVQLADHYQQNTPIGDGPVR (2849.087 Da) yields the spectra shown. Replacement of Met by Tfn results in a 36 amu mass increase.

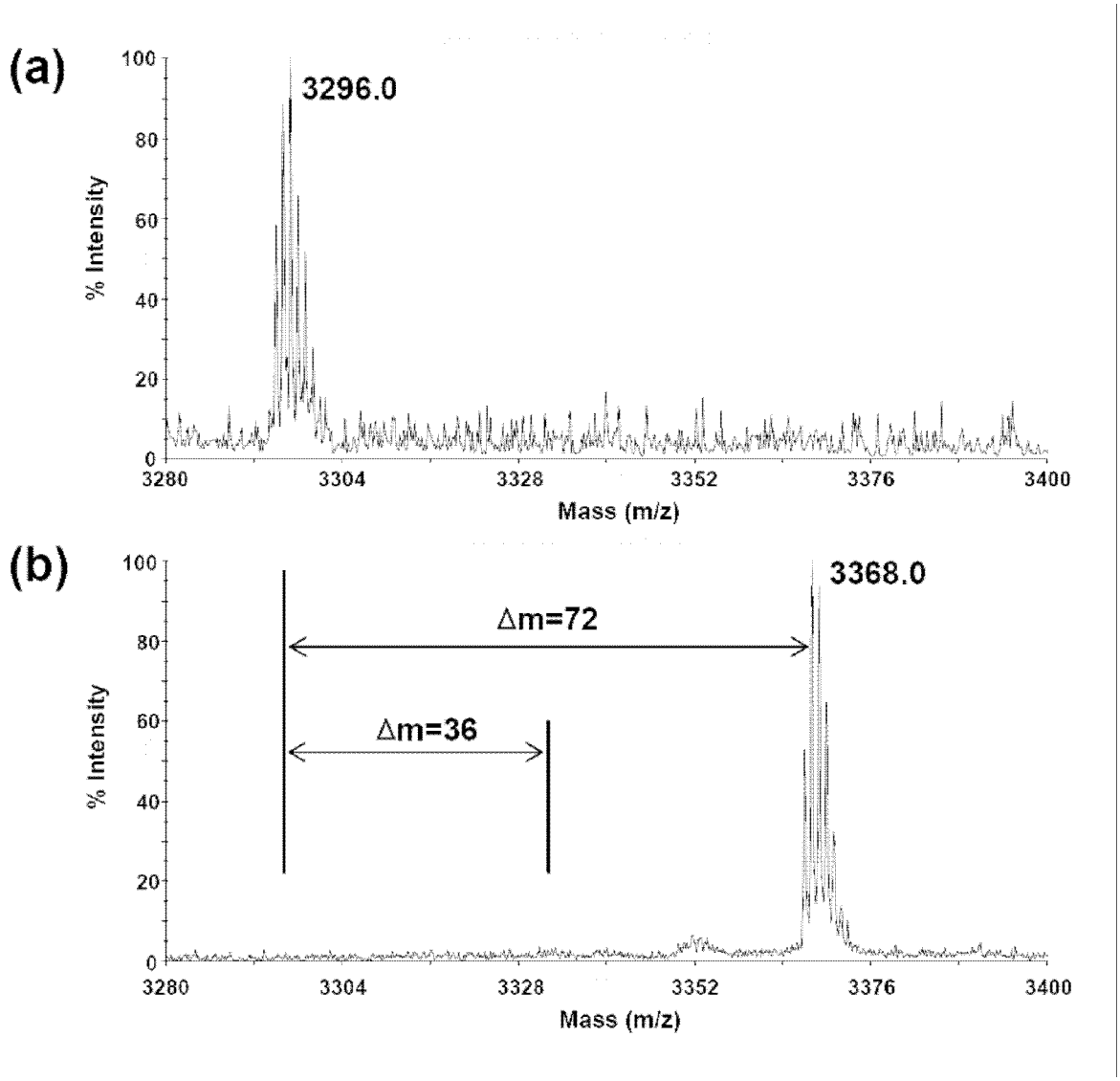

Figure S3. MALDI-MS analysis of mDHFR after trypsin digestion. The protein was expressed with 20 amino acids (a) or 19 amino acids plus 2 mM Tfn (b), and purified under denaturing conditions. A peptide fragment of sequence GSHHHHHHGSGIMVRPLNSIVAVSQNMGIGK (3295.750 Da) yields the spectra shown. Replacement of Met by Tfn results in a 36 amu mass increase.

Table S1. Amino acid sequences of wild-type GFP (WT GFP), L024_3-3, GFPrm, and GFPrm_AM[a]

	5									10									15									20									25									30	
WT GFP	M	S	K	G	E	E	L	F	T	G	V	V	P	I	L	V	E	L	D	G	D	V	N	G	H	K	F	S	V	S															
L024_3-3																														R																
GFPrm																														R																
GFRrm_AM																														R																

	35									40									45									50									55									60	
WT GFP	G	E	G	E	G	D	A	T	Y	G	K	L	T	L	K	F	I	C	T	T	G	K	L	P	V	P	W	P	T	L															
L024_3-3												I				L																															
GFPrm												I				L																															
GFPrm_AM												I				L																															

	65									70									75									80									85									90	
WT GFP	V	T	T	F	S	Y	G	V	Q	C	F	S	R	Y	P	D	H	M	K	Q	H	D	F	F	K	S	A	M	P	E															
L024_3-3				C	G							A								R																											
GFPrm				C	G							A						L		R									F																			
GFPrm_AM				C	G							A						L		R									F																			

	95									100									105									110									115									120	
WT GFP	G	Y	V	Q	E	R	T	I	F	F	K	D	D	G	N	Y	K	T	R	A	E	V	K	F	E	G	D	T	L	V															
L024_3-3									S						K	F													I																			
GFPrm									S						K	F													I																			
GFPrm_AM									S						K	F											M		I																			

	125									130									135									140									145									150	
WT GFP	N	R	I	E	L	K	G	I	D	F	K	E	D	G	N	I	L	G	H	K	L	E	Y	N	Y	N	S	H	N	V															
L024_3-3				K																									D																		
GFPrm				K																									D																		
GFPrm_AM				K							M																		D																		

	155								160								165								170								175								180
WT GFP	Y	I	M	A	D	K	Q	K	N	G	I	K	V	N	F	K	I	R	H	N	I	E	D	G	S	V	Q	L	A	D
L024_3-3			T						T				A								V										
GFPrm			T						T				A								V										
GFPrm_AM			T				M		T				A								V	M									

	185								190								195								200								205								210
WT GFP	H	Y	Q	Q	N	T	P	I	G	D	G	P	V	L	L	P	D	N	H	Y	L	S	T	Q	S	A	L	S	K	D
L024_3-3														R								L				V	I			
GFPrm														R								L				V	I			
GFPrm_AM														R								L				V	I			

	215								220								225								230								235				238
WT GFP	P	N	E	K	R	D	H	M	V	L	L	E	F	V	T	A	A	G	I	T	H	G	M	D	E	L	Y	K		
L024_3-3											H																				
GFPrm								A			H												I								
GFPrm_AM				M				A			H												I								

[a] A hexahistidine tag of sequence MRGSHHHHHHGS was appended to the N-terminus of each protein to enable purification by Ni-NTA chromatography.

Table S2. Sequence changes of 10 randomly selected clones from M02c_2. The single letters in parentheses represent the encoded amino acid.

WT	L13 CAG	A256 GCA	P257 CCG	Y260 TAC	H301 CAC
1	TCT (S)	GCG (A)	CCG (P)	TTG (L)	CTG (L)
2	AGT (S)	GCG (A)	CCG (P)	CTG (L)	CTT (L)
3	TCG (S)	GCG (A)	CCT (P)	CTT (L)	CTG (L)
4	AGT (S)	GCT (A)	CCT (P)	CTT (L)	CTG (L)
5	TCG (S)	GCT (A)	CCG (P)	CTT (L)	CTT (L)
6	TCT (S)	GCG (A)	CCT (P)	TTG (L)	CTG (L)
7	GCG (A)	GCG (A)	CCG (P)	CTT (L)	CTG (L)
8	AGT (S)	GCG (A)	CCG (P)	CTG (L)	CTT (L)
9	AGT (S)	GCA (A)[a]	CCG (P)	TTG (L)	TTG (L)
10	AGT (S)	GCT (A)	CCT (P)	CTG (L)	TTG (L)

[a] GCA codon for A256 of clone 9 might come from errors in synthesis of oligonucleotides.

Table S3. Kinetic parameters for activation of Met and Tfn by the M02c_2-8 MetRS variant.

Amino acid	k_{cat}, s^{-1}	K_m, μM	k_{cat}/K_m, $\mu M^{-1}s^{-1}$
Met	10.90 ± 0.24	5313 ± 109	2.1×10^{-3}
Tfn	3.85 ± 0.92	13233 ± 1789	2.9×10^{-4}